EGAD

HT1220 Report

RECORD INFORMATION

Gene ID:	1220
Sequence ID:	1220
Protein ID:	1220
Sequence name:	thrombospondin 1, alt. transcript 1
Genome:	nucleus
Taxon:	Homo sapiens
Locus:	1220
Common Name:	thrombospondin 1
Role ID:	40

Coding sequence length:	3513 nt
Transcript sequence length:	5722 nt
Expression data:	THC201673

ACCESSION DATA

HT1220 is derived from accessions(s):

SP:P07996 (THROMBOSPONDIN 1 PRECURSOR.)
GB:X04665 (Human mRNA for thrombospondin)
GB:X14787 (Human mRNA for thrombospondin)
GB:U12471 (thrombospondin-p50 {Homo sapiens})
GB:M99425 (Human thrombospondin mRNA, 3' end.)
PIR:G01478 (thrombospondin-p50 - human (fragment))
GB:U12471 (Human thrombospondin-1 gene, partial cds.)
GB:J04835 (Human thrombospondin gene, exons 1, 2 and 3.)
GB:M25631 (Homo sapiens (clone lambda-TS-33) thrombospondin (THBS) mRNA, 5' end.)

ALTERNATIVE SPLICE INFORMATION

Alternative splice forms for this gene:

HT3987 thrombospondin 1, alt. transcript 2

MAPPING DATA

GDB accession(s) for this gene:

GDB ID: Symbol
----

gdb:120438   THBS1

cDNA FEATURES

```
Feature                End 5  End 3
- - - - - - - - - - - - - - - - - - - - - - - - - - - - -
coding_seq             112    3624
3'UT                   3625   5722
spjunc_h               1235   1236
```

SEQUENCE

nucleotide:

```
ggacgcacaggcattccccgcgcccctccagccctcgccgccctcgccaccgctcccggc
cgccgcgctccggtacacacaggatccctgctgggcaccaacagctccaccatggggctg
gcctgggggactaggcgtcctgttcctgatgcatgtgtgtggcaccaaccgcattccagag
tctggcggagacaacagcgtgtttgacatcttttgaactcaccggggccgcccgcaagggg
tctgggcgccgactggtgaagggcccccgacccttccagcccagctttccgcatcgaggat
gccaacctgatcccccctgtgcctgatgacaagttccaagacctggtggatgctgtgcgg
gcagaaaagggtttcctccttctggcatccctgaggcagatgaagaagacccggggcacg
ctgctggccctggagcggaaagaccactctggccaggtcttcagcgtggtgtccaatggc
aaggcgggcaccctggacctcagcctgaccgtccaaggaaagcagcacgtggtgtctgtg
gaagaagctctcctggcaaccggccagtggaagagcatcaccctgtttgtgcaggaagac
agggcccagctgtacatcgactgtgaaaagatggagaatgctgagttggacgtccccatc
caaagcgtcttcaccagagacctggccagcatcgccagactccgcatcgcaaaggggggc
gtcaatgacaatttccaggggggtgctgcagaatgtgaggtttgtctttggaaccacacca
gaagacatcctcaggaacaaaggctgctccagctctaccagtgtcctcctcacccttgac
aacaacgtggtgaatggttccagccctgccatccgcactaactacattggccacaagaca
aaggacttgcaagccatctgcggcatctcctgtgatgagctgtccagcatggtcctggaa
ctcaggggcctgcgcaccattgtgaccacgctgcaggacagcatccgcaaagtgactgaa
gagaacaaagagttggccaatgagctgaggcggcctcccctatgctatcacaacggagtt
cagtacagaaataacgaggaatggactgttgatagctgcactgagtgtcactgtcagaac
tcagttaccatctgcaaaaaggtgtcctgccccatcatgcccctgctccaatgccacagtt
cctgatggagaatgctgtcctcgctgttggcccagcgactctgcggacgatggctggtct
ccatggtccgagtggacctcctgttctacgagctgtggcaatggaattcagcagcgcggc
cgctcctgcgatagcctcaacaaccgatgtgagggctcctcggtccagacacggacctgc
cacattcaggagtgtgacaaaagatttaaacaggatggtggctggagccactggtccccg
tggtcatcttgttctgtgacatgtggtgatggtgtgatcacaaggatccggctctgcaac
tctcccagcccccagatgaatgggaaaccctgtgaaggcgaagcgcgggagaccaaagcc
tgcaagaaagacgcctgccccatcaatggaggctggggtccttggtcaccatgggacatc
tgttctgtcacctgtggaggagggtacagaaacgtagtcgtctctgcaacaaccccgca
ccccagtttggaggcaaggactgcgttggtgatgtaacagaaaaccagatctgcaacaag
caggactgtccaattgatggatgcctgtccaatccctgctttgccggcgtgaagtgtact
agctaccctgatggcagctggaaatgtggtgcttgtcccccctggttacagtggaaatggc
atccagtgcacagatgttgatgagtgcaaagaagtgcctgatgcctgcttcaaccacaat
ggagagcaccggtgtgagaacacggaccccggctacaactgcctgccctgcccccacgc
ttcaccggctcacagcccttcggccagggtgtcgaacatgccacggccaacaaacaggtg
tgcaagcccccgtaacccctgcacggatgggacccacgactgcaacaagaacgccaagtgc
aactacctgggccactatagcgaccccatgtaccgctgcgagtgcaagcctggctacgct
ggcaatggcatcatctgcggggaggacacagacctggatggctggcccaatgagaacctg
gtgtgcgtggccaatgcgacttaccactgcaaaaaggataattgccccaaccttcccaac
tcagggcaggaagactatgacaaggatggaattggtgatgcctgtgatgatgacgatgac
aatgataaaattccagatgacagggacaactgtccattccattacaacccagctcagtat
gactatgacagagatgatgtgggagaccgctgtgacaactgtccctacaaccacaaccca
```

Figure 1B

```
gatcaggcagacacagacaacaatggggaaggagacgcctgtgctgcagacattgatgga
gacggtatcctcaatgaacgggacaactgccagtacgtctacaatgtggaccagagagac
actgatatggatgggggttggagatcagtgtgacaattgcccttggaacacaatccggat
cagctggactctgactcagaccgcattggagatacctgtgacaacaatcaggatattgat
gaagatggccaccagaacaatctggacaactgtccctatgtgcccaatgccaaccaggct
gaccatgacaaagatggcaagggagatgcctgtgaccacgatgatgacaacgatggcatt
cctgatgacaaggacaactgcagactcgtgcccaatcccgaccagaaggactctgacggc
gatggtcgaggtgatgcctgcaaagatgattttgaccatgacagtgtgccagacatcgat
gacatctgtcctgagaatgttgacatcagtgagaccgatttccgccgattccagatgatt
cctctggaccccaaagggacatcccaaaatgaccctaactgggttgtacgccatcagggt
aaagaactcgtccagactgtcaactgtgatcctggactcgctgtaggttatgatgagttt
aatgctgtggacttcagtggcaccttcttcatcaacaccgaaagggacgatgactatgct
ggatttgtctttggctaccagtccagcagccgctttttatgttgtgatgtggaagcaagtc
acccagtcctactgggacaccaaccccacgagggctcagggatactcgggcctttctgtg
aaagttgtaaactccaccacagggcctggcgagcacctgcggaacgccctgtggcacaca
ggaaacacccctggccaggtgcgcaccctgtggcatgaccctcgtcacataggctggaaa
gatttcaccgcctacagatggcgtctcagccacaggccaaagacgggtttcattagagtg
gtgatgtatgaagggaagaaaatcatggctgactcaggacccatctatgataaaacctat
gctggtggtagactagggttgtttgtcttctctcaagaaatggtgttcttctctgacctg
aaatacgaatgtagagatccctaatcatcaaattgttgattgaaagactgatcataaacc
aatgctggtattgcaccttctggaactatgggcttgagaaaacccccaggatcacttctc
cttggcttccttctttttctgtgcttgcatcagtgtggactcctagaacgtgcgacctgcc
tcaagaaatgcagttttcaaaaacagactcatcagcattcagcctccaatgaataagac
atcttccaagcatatataaacaattgctttggtttccttttgaaaaagcatctacttgcttc
agttgggaaggtgcccattccactctgcctttgtcacagagcagggtgctattgtgaggc
catctctgagcagtggactcaaaagcattttcaggcatgtcagagaagggaggactcact
agaattagcaaacaaaccaccctgacatcctccttcaggaacacggggagcagaggcca
aagcactaaggggagggcgcatacccgagacgattgtatgaagaaaatatggaggaactg
ttacatgttcggtactaagtcattttcagggggattgaaagactattgctggatttcatga
tgctgactggcgttagctgattaacccatgtaaataggcacttaaatagaagcaggaaag
ggagacaaagactggcttctggacttcctccctgatccccacccttactcatcaccttgc
agtggccagaattagggaatcagaatcaaaccagtgtaaggcagtgctggctgccattgc
ctggtcacattgaaattggtggcttcattctagatgtagcttgtgcagatgtagcaggaa
aataggaaaacctaccatctcagtgagcaccagctgcctcccaaaggaggggcagccgtg
cttatatttttatggttacaatggcacaaaattattatcaacctaactaaaacattcctt
ttctcttttttccgtaattactaggtagttttctaattctctcttttggaagtatgattt
ttttaaagtctttacgatgtaaaatatttattttttacttattctggaagatctggctga
aggattattcatggaacaggaagaagcgtaaagactatccatgtcatctttgttgagagt
cttcgtgactgtaagattgtaaatacagattatttattaactctgttctgcctggaaatt
taggcttcatacggaaagtgtttgagagcaagtagttgacatttatcagcaaatctcttg
caagaacagcacaaggaaaatcagtctaataagctgctctgccccttgtgctcagagtgg
atgttatgggattcctttttttctctgtttttatctttttcaagtggaattagttggttatcc
atttgcaaatgttttaaattgcaaagaaagccatgaggtcttcaatactgttttacccca
tcccttgtgcatatttccaggagagaaggaaagcatatacacttttttcttttcattttccc
aaaagagaaaaaaatgacaaaaggtgaaacttacatacaaatattacctcatttgttgtg
tgactgagtaaagaatttttggatcaagcggaaagagtttaagtgtctaacaaacttaaa
gctactgtagtacctaaaaagtcagtgttgtacatagcataaaaactctgcagagaagta
ttcccaataaggaaatagcattgaaatgttaaatacaatttctgaaagttatgttttttt
tctatcatctggtataccattgctttattttttataaattatttttctcattgccattggaa
tagaatattcagattgtgtagatatgctatttaaataatttatcaggaaatactgcctgt
agagttagtatttctattttatataatgtttgcacactgaattgaagaattgttggttt
tttctttttttgttttttttttttttttttttttttgcttttgacctcccatttttta
ctatttgccaatacctttttctaggaatgtgctttttttttgtacacattttatccattt
tacattctaaagcagtgtaagttgtatattactgtttcttatgtacaaggaacaacaata
aatcatatggaaatttatattt
```

protein:

MGLAWGLGVLFLMHVCGTNRIPESGGDNSVFDIFELTGAARKGSGRRLVKGPDPSSPAFR

Figure 1C

IEDANLIPPVPDDKFQDLVDAVRAEKGFLLLASLRQMKKTRGTLLALERKDHSGQVFSVV
SNGKAGTLDLSLTVQGKQHVVSVEEALLATGQWKSITLFVQEDRAQLYIDCEKMENAELD
VPIQSVFTRDLASIARLRIAKGGVNDNFQGVLQNVRFVFGTTPEDILRNKGCSSSTSVLL
TLDNNVVNGSSPAIRTNYIGHKTKDLQAICGISCDELSSMVLELRGLRTIVTTLQDSIRK
VTEENKELANELRRPPLCYHNGVQYRNNEEWTVDSCTECHCQNSVTICKKVSCPIMPCSN
ATVPDGECCPRCWPSDSADDGWSPWSEWTSCSTSCGNGIQQRGRSCDSLNNRCEGSSVQT
RTCHIQECDKRFKQDGGWSHWSPWSSCSVTCGDGVITRIRLCNSPSPQMNGKPCEGEARE
TKACKKDACPINGGWGPWSPWDICSVTCGGGVQKRSRLCNNPAPQFGGKDCVGDVTENQI
CNKQDCPIDGCLSNPCFAGVKCTSYPDGSWKCGACPPGYSGNGIQCTDVDECKEVPDACF
NHNGEHRCENTDPGYNCLPCPPRFTGSQPFGQGVEHATANKQVCKPRNPCTDGTHDCNKN
AKCNYLGHYSDPMYRCECKPGYAGNGIICGEDTDLDGWPNENLVCVANATYHCKKDNCPN
LPNSGQEDYDKDGIGDACDDDDDNDKIPDDRDNCPFHYNPAQYDYDRDDVGDRCDNCPYN
HNPDQADTDNNGEGDACAADIDGDGILNERDNCQYVYNVDQRDTDMDGVGDQCDNCPLEH
NPDQLDSDSDRIGDTCDNNQDIDEDGHQNNLDNCPYVPNANQADHDKDGKGDACDHDDDN
DGIPDDKDNCRLVPNPDQKDSDGDGRGDACKDDFDHDSVPDIDDICPENVDISETDFRRF
QMIPLDPKGTSQNDPNWVVRHQGKELVQTVNCDPGLAVGYDEFNAVDFSGTFFINTERDD
DYAGFVFGYQSSSRFYVVMWKQVTQSYWDTNPTRAQGYSGLSVKVVNSTTGPGEHLRNAL
WHTGNTPGQVRTLWHDPRHIGWKDFTAYRWRLSHRPKTGFIRVVMYEGKKIMADSGPIYD
KTYAGGRLGLFVFSQEMVFFSDLKYECRDP

Figure 1D
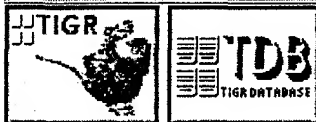
HT2143 Report

RECORD INFORMATION

Gene ID:	2081
Sequence ID:	2143
Protein ID:	2125
Sequence name:	thrombospondin 4
Genome:	nucleus
Taxon:	Homo sapiens
Locus:	2081
Common Name:	thrombospondin 4
Role ID:	40

Coding sequence length:	2886 nt
Transcript sequence length:	3074 nt
Expression data:	THC168897

ACCESSION DATA

HT2143 is derived from accessions(s):

SP:P35443 (THROMBOSPONDIN 4 PRECURSOR.)
GB:Z19585 (thrombospondin-4 {Homo sapiens})
GB:Z19585 (H.sapiens mRNA for thrombospondin-4)
PIR:A55710 (thrombospondin 4 precursor - human)

cDNA FEATURES

Feature	End 5	End 3
coding_seq	28	2913
3'UT	2914	3074

SEQUENCE

nucleotide:

```
gaattccggggagcaggaagagccaacatgctggccccgcgcggagccgccgtcctcctg
ctgcacctggtcctgcagcggtggctagcggcaggcgcccaggccacccccaggtctttt
gaccttctcccatcttccagtcagaggctaaacccaggcgctctgctgccagtcctgaca
gaccccgccctgaatgatctctatgtgatttccaccttcaagctgcagactaaaagttca
gccaccatcttcggtctttactcttcaactgacaacagtaaatattttgaatttactgtg
atgggacgcttaagcaaagccatcctccgttacctgaagaacgatgggaaggtgcatttg
```

Figure 2A

```
gtggttttcaacaacctgcagctggcagacggaaggcggcacaggatcctcctgaggctg
agcaatttgcagcgaggggccggctccctagagctctacctggactgcatccaggtggat
tccgttcacaatctccccagggcctttgctggcccctcccagaaacctgagaccattgaa
ttgaggactttccagaggaagccacaggacttcttggaagagctgaagctggtggtgaga
ggctcactgttccaggtggccagcctgcaagactgcttcctgcagcagagtgagccactg
gctgccacaggcacaggggactttaaccggcagttcttgggtcaaatgacacaattaaac
caactcctgggagaggtgaaggaccttctgagacagcaggttaaggaaacatcatttttg
cgaaacaccatagctgaatgccaggcttgcggtcctctcaagtttcagtctccgacccca
agcacggtggtcgccccggctccccctgcaccgccaacacgcccacctcgtcggtgtgac
tccaacccatgtttccgaggtgtccaatgtaccgacagtagagatggcttccagtgtggg
ccctgccccgagggctacacaggaaacgggatcacctgtattgatgttgatgagtgcaaa
taccatccctgctacccgggcgtgcactgcataaatttgtctcctggcttcagatgtgac
gcctgcccagtgggcttcacagggcccatggtgcagggtgttgggatcagttttgccaag
tcaaacaagcaggtctgcactgacattgatgagtgtcgaaatggagcgtgcgttcccaac
tcgatctgcgttaatactttgggatcttaccgctgtgggccttgtaagccggggtatact
ggtgatcagataaggggatgcaaagtggaaagaaactgcagaaacccagagctgaaccct
tgcagtgtgaatgcccagtgcattgaagagaggcaggggatgtgacatgtgtgtgtgga
gtcggttgggctggagatggctatatctgtggaaaggatgtggacatcgacagttacccc
gacgaagaactgccatgctctgccaggaactgtaaaaaggacaactgcaaatatgtgcca
aattctggccaagaagatgcagacagagatggcattggcgacgcttgtgacgaggatgct
gacggagatgggatcctgaatgagcaggataactgtgtcctgattcataatgtggaccaa
aggaacagcgataaagatatctttggggatgcctgtgataactgcctgagtgtcttaaat
aacgaccagaaagacaccgatggggatggaagaggagatgcctgtgatgatgacatggat
ggagatggaataaaaaacattctggacaactgcccaaaatttcccaatcgtgaccaacgg
gacaaggatggtgatggtgtggggatgcctgtgacagttgtcctgatgtcagcaaccct
aaccagtctgatgtggataatgatctggttggggactcctgtgacaccaatcaggacagt
gatggagatgggcaccaggacagcacagacaactgccccaccgtcattaacagtgcccag
ctggacaccgataaggatggaattggtgacgagtgtgatgatgatgatgacaatgatggt
atcccagacctggtgccccctggaccagacaactgccggctggtccccaacccagcccag
gaggatagcaacagcgacggagtgggagacatctgtgagtctgactttgaccaggaccag
gtcatcgatcggatcgacgtctgcccagagaacgcagaggtcaccctgaccgacttcagg
gcttaccagaccgtgggcctggatcctgaaggggatgcccagatcgatcccaactgggtg
gtcctgaaccagggcatggagattgtacagaccatgaacagtgatcctggcctggcagtg
gggtacacagctttttaatggagttgacttcgaagggaccttccatgtgaatacccagaca
gatgatgactatgcaggctttatctttggctaccaagatagctccagcttctacgtggtc
atgtggaagcagacggagcagacatattggcaagccaccccattccgagcagttgcagaa
cctggcattcagctcaaggctgtgaagtctaagacaggtccaggggagcatctccggaac
tccctgtggcacacggggacaccagtgaccaggtcaggctgctgtggaaggactccagg
aatgtgggctggaaggacaaggtgtcctaccgctggttcctacagcacaggccccaggtg
ggctacatcagggtacgatttttatgaaggctctgagttggtggctgactctggcgtcacc
atagacaccacaatgcgtggaggccgacttggcgttttctgcttctctcaagaaaacatc
atctggtccaacctcaagtatcgctgcaatgacaccatccctgaggacttccaagagttt
caaacccagaatttcgaccgcttcgataattaaaccaaggaagcaatctgtaactgcttt
tcggaacactaaaaccatatatattttaacttcaattttctttagcttttaccaacccaa
atatatcaaaacgttttatgtgaatgtggcaataaaggagaagagatcatttttaaaaaa
aaaaaaaaaaaaaa
```

protein:

```
MLAPRGAAVLLLHLVLQRWLAAGAQATPQVFDLLPSSSQRLNPGALLPVLTDPALNDLYV
ISTFKLQTKSSATIFGLYSSTDNSKYFEFTVMGRLSKAILRYLKNDGKVHLVVFNNLQLA
DGRRHRILLRLSNLQRGAGSLELYLDCIQVDSVHNLPRAFAGPSQKPETIELRTFQRKPQ
DFLEELKLVVRGSLFQVASLQDCFLQQSEPLAATGTGDFNRQFLGQMTQLNQLLGEVKDL
LRQQVKETSFLRNTIAECQACGPLKFQSPTPSTVVAPAPPAPPTRPPRRCDSNPCFRGVQ
CTDSRDGFQCGPCPEGYTGNGITCIDVDECKYHPCYPGVHCINLSPGFRCDACPVGFTGP
MVQGVGISFAKSNKQVCTDIDECRNGACVPNSICVNTLGSYRCGPCKPGYTGDQIRGCKV
ERNCRNPELNPCSVNAQCIEERQGDVTCVCGVGWAGDGYICGKDVDIDSYPDEELPCSAR
NCKKDNCKYVPNSGQEDADRDGIGDACDEDADGDGILNEQDNCVLIHNVDQRNSDKDIFG
DACDNCLSVLNNDQKDTDGDGRGDACDDDMDGDGIKNILDNCPKFPNRDQRDKDGDGVGD
```

Figure 2B

ACDSCPDVSNPNQSDVDNDLVGDSCDTNQDSDGDGHQDSTDNCPTVINSAQLDTDKDGIG
DECDDDDDNDGIPDLVPPGPDNCRLVPNPAQEDSNSDGVGDICESDFDQDQVIDRIDVCP
ENAEVTLTDFRAYQTVGLDPEGDAQIDPNWVVLNQGMEIVQTMNSDPGLAVGYTAFNGVD
FEGTFHVNTQTDDDYAGFIFGYQDSSSFYVVMWKQTEQTYWQATPFRAVAEPGIQLKAVK
SKTGPGEHLRNSLWHTGDTSDQVRLLWKDSRNVGWKDKVSYRWFLQHRPQVGYIRVRFYE
GSELVADSGVTIDTTMRGGRLGVFCFSQENIIWSNLKYRCNDTIPEDFQEFQTQNFDRFD
N

TIGR

TDB
TIGR DATABASE

Figure 2C

Poly ID	Sequence ID	Position	Gene Description	Flanking Seq	Mutation T	Ref NT	Alt NT	Ref AA	Alt AA
G3344	HTHT1220_mRNA	2210	THBS1, thrombospondin 1	TGGATGGCTGGCCCAIA/GITGAGAACCTGGTGTG	Missense	A	G	N	S
G3355u2	HTHT2143_mRNA	1186	THBS4, thrombospondin 4	GAGTGTCGAAATGGA[G/C]CGTGCGTTCCCAACT	Missense	G	C	A	P

Figure 3